

Kepler - Bug #4161

Develop kepler workflow execution engine

06/15/2009 03:06 PM - Jing Tao

<b>Status:</b>	Resolved	<b>Start date:</b>	06/15/2009
<b>Priority:</b>	Immediate	<b>Due date:</b>	
<b>Assignee:</b>	Jing Tao	<b>% Done:</b>	0%
<b>Category:</b>	general	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	wrp-modules-2.0.0	<b>Spent time:</b>	0.00 hour
<b>Bugzilla-Id:</b>	4161		
<b>Description</b> Here is the an email from matt:  I think the next stage would be to have you work on putting together an execution engine that can, given a schedule of a workflow to be run, can be called to retrieve the workflow KAR from metacat, set up a proper staging directory for execution, and run kepler, making sure that output and errors are handled properly. On the output side, this means programatically calling the run manager that Derik and Aaron are making to create a kar file that contains the results, and then publish that to metacat (Kepler also should have an API for this when Derik and Aaron are done).  We decided to use web service as the execution engine. The API will be:  1. execute(workflowID)::runID 2. getStatus(runID)::status  status includes: SCHEDULED EXECUTING SUCCEDED FAILED NOTFOUND			
<b>Related issues:</b>			
Blocks Kepler - Bug #4169: Build "server-side" installation of Kepler		Resolved	06/18/2009
Blocks Kepler - Bug #4180: Kepler command-line execution with cache		Resolved	06/23/2009

History

#1 - 06/16/2009 03:09 PM - Jing Tao

In execute(workflowID) method, we need to get the kar file from metacat first.  
To my understanding. The workflowID will be a lsid format. Do we have any mechanism to get the file from remote repository base on given lsid?  
  
If we don't have this mechanism, we may use metacat id too. So the API will be:  
execute(workflowID, metacatID)::runID  
  
Any suggestion and comment will be appreciated.

#2 - 06/16/2009 03:11 PM - Jing Tao

Of course, adding metacatID is a short term solution.

#3 - 06/16/2009 03:25 PM - Jing Tao

Here is matt's comment:  
<matt> hey jing  
<matt> i saw your bugzilla note about retrieving KAR files by LSID  
<jing> hi, matt  
<matt> The LSID authority service should let you get the KAR file using the getData(LSID) method  
<matt> however, it might be better to not rely on that  
<matt> and instead, recognize that the ID is an LSID, extract out the metacat identifier part, then make a standard call to the ecogrid get method to get the kar

<matt> do you think that would work?  
<jing> if we make sure metacat identifier is part of LSID, it should work.  
<matt> it is, in that the metacat id is the namespace:object:rev part of the lsid  
<matt> so there is a 1:1 mapping  
<jing> sounds good.

#### #4 - 06/18/2009 09:08 AM - ben leinfelder

After discussion with Jing and Derik yesterday:

- 1-we need to make sure we have a server-side version of Kepler that includes the WRP suite.
- 2-we need to verify that provenance and reporting features work GUI-less
- 3-we need to make some actions automated: packaging a KAR after execution; uploading that KAR.

will [#3](#) be handled by Jing or by the reporting module?

#### #5 - 06/18/2009 11:11 AM - ben leinfelder

In the client-side GUI version of Kepler we open/run workflows and reports like so:

- 1-Open KAR (contains the workflow and the report layout)
- 2-Run the workflow (which generates the report when completed and writes it to provenance)
- 3-Export the run as KAR (contains the workflow that was run, the report layout, the report instance (xml) and a pdf of the report instance).

The execution engine needs to open the workflow from the KAR so that the report layout is available. I believe the current plan *only* involves the workflow LSID and does not address any report layout issues.

I'd like to see an execution method that can take a KAR LSID (that contains a workflow and a report layout). It would then perform the three steps above.

Note that the KAREntryHandlers take care of the opening/initializing steps for reporting to work - so that's good (and already built in)!

#### #6 - 07/07/2009 09:51 AM - ben leinfelder

we've had a couple successful "round trip" runs:

```
upload a workflow KAR (indus)
schedule execution (indus)
execute wf (chico)
archive results (chico>indus)
-view results (indus)
```

Now we've got to resolve the LSID-changing problem that is happening when we open a KAR with a workflow (bug [#4224](#)). Then we should be able to run another end-to-end test

#### #7 - 10/16/2009 11:19 AM - ben leinfelder

re-targeting to wrp module, although it may be a 2.0 bug to extend kepler so that it can receive user credentials when running from the command line.

#### #8 - 11/18/2009 10:07 AM - ben leinfelder

In terms of the credentials - we can now include them in the configuration xml file that is used by the command line version of Kepler:  
ConfigNoGUIWithCache.xml

```
<property name="_domain" class="ptolemy.kernel.util.StringAttribute" value="KNB,DEV" />
<property name="_username" class="ptolemy.kernel.util.StringAttribute" value="uid=kepler,o=unaffiliated,dc=ecoinformatics,dc=org" />
<property name="_password" class="ptolemy.kernel.util.StringAttribute" value="kepler" />
```

but with the correct user for the TPC group

#### #9 - 01/05/2010 11:40 AM - ben leinfelder

script-based execution still works. There's a slight problem on the Metacat side for the first MOML DTD...but that's on it's way to being resolved. The report for my test was blank, however...that's a problem.

#### #10 - 01/05/2010 04:08 PM - ben leinfelder

stylesheet issue on the Metacat side of things -- now fixed.

#### #11 - 01/06/2010 11:36 AM - ben leinfelder

installed on my local system - did a roundtrip:

```
-upload workflow+report
-schedule
-wait
-view PDF report when complete
```

There are a few minor Metacat updates to be committed, but otherwise things are still working.

**#12 - 02/04/2010 12:49 PM - ben leinfelder**

Need to move the authentication credentials from the old config system (ptolemy-based) to the new configuration.xml system.  
Seems as if we will not be passing these params on the commandline, so a purely configuration-based approach will suffice for now.

**#13 - 03/30/2010 06:52 PM - Derik Barseghian**

Hey Jing and/or Ben, could you give an update on this bug?

**#14 - 03/31/2010 07:41 AM - ben leinfelder**

I know in early february I was able to use the commandline to execute a KAR (after some tweaks) and then Josep was unknowingly helping us make enhancements.

As long as the commandline execution works, I think the webservice-based execution is pretty straightforward.

We should update the dev (metacat) server to have a more recent WRP build of Kepler and make sure the TPC scheduling system is configured to use it.

**#15 - 03/31/2010 09:48 AM - Jing Tao**

Early of March, I installed an execution engine on my local machine. I finished a full a cycle - uploaded a kar file from kepler to the server, searched the kar file on the web site, scheduled a job and the job was executed successfully.

So I think the engine works. But it may still need some improvement.

**#16 - 04/19/2010 01:15 PM - Jing Tao**

so far the engine works. I am closing the bug.

**#17 - 04/19/2010 01:16 PM - Jing Tao**

closing the bug.

**#18 - 03/27/2013 02:25 PM - Redmine Admin**

Original Bugzilla ID was 4161